Model-guided \textit{in silico} evaluation of the effect of pfl gene knockouts on the production of D-lactate by \textit{Escherichia coli} using the OptFlux software platform

Bashir Sajo Mienda* and Mohd Shahir Shamsir

Bioinformatics Research Group (BIRG), Biosciences and Health Sciences Department, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, Skudai 81310 Johor Bahru. Malaysia

*Corresponding author: bsmienda@gmail.com

Abstract

The increase availability of genome scale metabolic models of \textit{Escherichia coli} and computational successes is revolutionizing the field of metabolic engineering and synthetic microbiology. \textit{E. coli} has been experimentally established to produce D-lactate under micro-aerobic conditions when pyruvate formate lyase (PFL) genes are knocked out. However, investigation on the \textit{in silico} prediction and/or evaluation of the effect of PFL genes knockout on the production of D-lactate using \textit{E. coli} genome scale metabolic model with Regulatory on / off Minimization (ROOM) under the OptFlux software platform remained underexplored. We demonstrated the metabolic engineering strategies using the OptFlux software platform by gene knockout simulation of pflA/b0902, pflB/b0903, pflC/b3952 and pflD/b3951 have been predicted to increase D-lactate production in \textit{E. coli} and hence maintaining a growth rate that is 96% of the wild-type model. The deletions of the PFL genes have been established to increase D-lactate production in \textit{E. coli}. The results obtained in this study are in agreement with the previously established experimental studies. These findings suggests that the OptFlux software platform using ROOM as the simulation algorithm, can prospectively and effectively predict future metabolic engineering targets for increased D-lactate production in \textit{E. coli} or other microbial chemical syntheses.

Keywords: \textit{Escherichia coli} model, OptFlux software, D-lactate, metabolic engineering, gene knockout simulation

Introduction

Lactic acid (lactate) and some its derivatives have many application in the food, pharmaceutical and polymer industries. \textit{Escherichia coli} as a well-engineered chassis host used to have a significant metabolic shift between aerobic and anaerobic conditions, as a result \textit{E. coli} produces a mixture of of organic acids such as formate, acetate, D-Lactate, succinate and ethanol under oxygen limited conditions without other electron acceptors (Zhu and Shimizu, 2004). D-lactate was known to be produced by \textit{E. coli} under micro aerobic conditions when pyruvate formate lyase genes are knocked out (Zhu and Shimizu, 2004). D-lactate industrial production is beset with a number of challenges, which entails improved \textit{E. coli} strain with increase production potential. Metabolic engineering as a discipline has been and will continue to provide engineered \textit{E. coli strains} with great potential for industrial production of D-lactate. Although several laboratories have reported alternative biocatalysts (Zhu and Shimizu, 2004; Mazumdar et al., 2010; Okano et al., 2010), many of which constitutes engineered \textit{E. coli strains} that produce D- or L-lactate (Zhou et al., 2003a; Zhou et al., 2003b; Zhu and Shimizu, 2004; Zhu et al., 2007). The identification of suitable engineering targets using \textit{in silico} biotechnology
approaches would offer an easy solution to guide future systems metabolic engineering for improved D-lactate production in *E. coli*.

The availability of genome scale models, particularly of *E. coli* (Feist et al., 2007; Orth et al., 2011) and the use of *in silico* metabolic engineering software platform to predict and identify metabolic gene knockouts targets have received remarkable attention in recent years (Mienda and Shamsir, 2014; Mienda et al., 2014a; Mienda et al., 2014b). One of the notable examples are the use of the OptFlux software to predict metabolic engineering interventions using *E. coli* genome scale as described previously (Rocha et al., 2010; Mienda et al., 2014b). OptKnock, is another software that was proposed to predict metabolic engineering for lactic acid production in *E. coli* (Fong et al., 2005; Mienda et al., 2014a). In addition, OptGene, was reported among other alternative methods that is capable of *in silico* gene knockout in genome scale metabolic models (Rocha et al., 2008). A peculiar drawback of the use of OptKnock and OptGene in metabolic engineering interventions is that they only use metabolic information, determining sets of reactions to be eliminated from the metabolic models, instead of sets of genes to knockout, which is the real purpose (Vilaça et al., 2011; Mienda et al., 2014a). Therefore, in order to have an appropriate mutant in the experimental setting, one needs to determine which sets of genes can lead to the elimination of a given set of reactions (Vilaça et al., 2011; Mienda et al., 2014a) This is because the rule-1 gene: 1 enzyme: 1 reaction-was not universal (Mienda et al., 2014a). However, there are many exceptions, such as isoenzymes, protein complexes, or enzymes that catalyze several reactions (Vilaça et al., 2011; Mienda et al., 2014a).

On the other hand, *E. coli* systems metabolic engineering could be synergistically combine with computational tools and synthetic biology to offer promising solutions for industrial chemical productions. An example is seen in a computational tool/software interface for *in silico* metabolic engineering interventions and constraint based modelling called OptFlux (Rocha et al., 2010). The OptFlux software platform uses genome scale metabolic models of *E. coli* to predict the phenotype simulation of both the wild-type and the mutant strain using the method of flux balance analysis (FBA) (Rocha et al., 2010; Mienda and Shamsir, 2014; Mienda et al., 2014a; Mienda et al., 2014b). The software has a peculiar feature of plug-in architecture, where an algorithm such as Regulatory on/off Minimization of Metabolic flux changes (ROOM) can be used to introduce genetic perturbations in the *E. coli* genome scale model, there by paving ways for model-guide experimental inquiry and/or understanding novel biological insight on the behavior of the mutant model after gene knockout (Shlomi et al., 2005; Rocha et al., 2010; Mienda et al., 2014a; Monk and Palsson, 2014). Computational tool such as OptKnock was reported to be a reference algorithm for studying a number of metabolic gene knockout using *E. coli* genome scale model using bi-level optimization approach for the D-lactate production, but it does not allow non-linear objective function and need a considerable to compute a solution (Fong et al., 2005). We previously reiterated how computational breakthroughs can help to redesign microbial chassis host for robust bioethanol production (Mienda and Shamsir, 2013). *In silico* metabolic engineering interventions for increased ethanol production from glucose and gluconate using the OptFlux software...
platform was reported (Mienda et al., 2014b). In a similar study, enhanced ethanol production by model-guided in silico metabolic engineering in *E. coli* was predicted using glycerol and xylose as the main solitary carbon sources (Mienda and Shamsir, 2014). Furthermore, D-lactate production from glycerol was predicted using *E. coli* genome-scale model with the OptFlux software platform (Mienda et al., 2014a). Similar software called, MetaFluxNet was used to investigate in silico metabolic engineering targets by comparative genome approach by Lee and co-workers (Lee et al., 2003; Lee et al., 2005) to increase succinate production in *E. coli* (Lee et al., 2005). Nevertheless, we used the OptFlux software platform, with a more advanced *in silico* metabolic engineering capabilities and peculiar plug-in architecture in current work. To the best of our knowledge, we report for the first time the implementation of this software to predict gene knockout in *E. coli* iJ01366 (Orth et al., 2011). We used ROOM as the algorithm for simulation to predict whole cell’s post perturbation behavior after gene knockout *(pflA/b0902, pflB/b0903, pflC/b3952 and pflD/b3951)* in relation to D-lactate production.

**Materials and Methods**

**Model**

The metabolic reconstruction of *Escherichia coli* iJ01366 (Orth et al., 2011) was used as a model for all the wild-type and mutant strains described in in this study. The model was previously tested and validated against experimental data, and was shown to be capable of predicting accurate growth rates, metabolite excretion rates, and a growth phenotypes on a number of substrates and genetic conditions (Feist et al., 2007; Feist et al., 2010; McCloskey et al., 2013). The substrates used in this study are glucose unless otherwise stated.

**Flux Balance analysis**

OptFlux software, as an open source platform www.optflux.org (Rocha et al., 2010) and a reference computational tool for metabolic engineering was used for the flux balance analysis (FBA). Regulatory on/off minimization of metabolic flux changes after genetic perturbations (ROOM) (Shlomi et al., 2005) was used as a simulation method for gene knockouts, and it was implemented using the Java programming within the framework of the OptFlux as described elsewhere (Rocha et al., 2010; Mienda et al., 2014a). All simulation of mutant strains and wild-type models were performed using the OptFluxv3.06. The chosen solitary carbon source is glucose, and the uptake rate of the carbon source was constrained to a maximum of 20 mmolgDW⁻¹h⁻¹. The oxygen uptake rate was considered to be 5 mmolgDW⁻¹h⁻¹ as the simulation condition was micro-aerobic for fermentative production of D-lactate. These values were chosen based on slightly close experimental observation of micro-aerobic and anaerobic growth of *E. coli* (Varma et al., 1993; Edwards et al., 2001; Fischer et al., 2004).

**Gene knockout under the OptFlux software platform**

Gene knockout simulation was conducted under the OptFlux software platform using ROOM (Shlomi et al., 2005) as simulation method. Flux balance analysis (FBA) was used for simulation of the wild-type model, which predicts metabolic flux distributions at steady state by using linear programming, while ROOM uses mixed integer linear programing (MILP) to find flux distribution that predicts the same constraint as FBA while minimizing the number of significant flux changes (Shlomi et al., 2005). The algorithm accounted only for flux changes (0.001 flux prediction) that is considered significant (Shlomi et al., 2005). This is because inherent noise some
time exists in biological systems and by using small flux changes, reduced running time is achieved as described in their original documentation (Shlomi et al., 2005). The wild-type model obtained from the Biomodels database (Le Novere et al., 2006), constructed by Orth (Orth et al., 2011) was designated as WT (Orth Model) and the mutant models/strains with pyruvate formate lyase single gene knockouts were designated as pflA, pflB, pflC and pflD strains respectively. The in silico gene knockouts simulation were run to completion using ROOM, as previously described in their original documentation (Shlomi et al., 2005).

**Results and Discussion**

Metabolism of *E. coli* experienced significant changes between aerobic and anaerobic conditions. Under oxygen limited conditions *E. coli* produces a mixture of organic acids such as formate, acetate, D-lactate, succinate and ethanol without other electron acceptors pyruvate is mainly assimilated via pyruvate formate lyase (PFL) and form formic acid and acetyl-CoA (AcCoA) (Zhu and Shimizu, 2004). It was previously established that knocking out of PFL genes (pflA, pflB, pflC and pflD) triggered a metabolic turn over towards the production of D-lactate under micro-aerobic conditions (Zhu and Shimizu, 2004). The in silico results of D-lactate production from glucose as substrate was indicated in table 1 and figure 1. It has been shown that the wild-type model (WT Orth) showed no D-lactate production when oxygen limited (micro-aerobic) condition was used for the simulation, indicating clearly that under micro-aerobic conditions, *E. coli* does not naturally produce D-lactate. On one hand the deletion of PFL genes in the *E. coli* central metabolism indicate a positive D-lactate production under micro-aerobic condition with a growth rate that is 96% of the wild-type model (Table 1). While on the other hand, no D-lactate was produced when the wild-type model was simulated under the same condition. These findings are in agreement with previously established experimental studies reported elsewhere (Zhu and Shimizu, 2004), where *pflA, pflB, pflC and pflD* were knocked out under micro-aerobic conditions to achieve a desired D-lactate production in *E. coli*.

**Fig. 1.** The Growth rates and D-lactate productivity of the wild-type and mutant *E. coli* models constructed in this study

It was previously reported that the conversion of one molecule of pyruvate to lactate requires one molecule of NADH and ethanol formation on the other hand needs two molecules of NADH. Alteration in the production of each metabolite, *E. coli* can modulate its metabolism fermentatively to grow on a number of substrates (Zhu and Shimizu, 2004). Alternative pathways could be activated to direct the carbon fluxes through the role of some key metabolites such as energy and electron donors as well as the metabolites at the important branch points such as pyruvate and AcCoA (zu Berstenhorst et al., 2009) (Fig. 2). In addition, it was previously established and known that mutation in the specific metabolic pathways such as those reported in this study, significantly affect the overall fermentation characteristics. Pyruvate is assimilated through PFL under micro-aerobic conditions (Svetlana...
Alexeeva and Mattos, 2000). Therefore, knocking out the genes that codes for PFL, lead to a significant metabolic interventions that results in increase or decrease production of a particular compound such as D-lactate for instance.

Table 1. *E. coli* strain design properties on glucose under the OptFlux software platform

<table>
<thead>
<tr>
<th><em>E. coli</em> Knock out genes/strains</th>
<th>Growth rate (h⁻¹)</th>
<th>% Biomass</th>
<th>D-lactate (mmolgDW⁻¹h⁻¹)</th>
<th>Acetate (mmolgDW⁻¹h⁻¹)</th>
<th>%</th>
<th>Ethanol (mmolgDW⁻¹h⁻¹)</th>
<th>Formate (mmolgDW⁻¹h⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT (orth Model)</td>
<td>0.77</td>
<td>100</td>
<td>0.00</td>
<td>19.35</td>
<td>100</td>
<td>9.15</td>
<td>31.48</td>
</tr>
<tr>
<td>pflA/b0902</td>
<td>0.75</td>
<td>96.86</td>
<td>0.001</td>
<td>18.79</td>
<td>97.1</td>
<td>8.88</td>
<td>30.57</td>
</tr>
<tr>
<td>pflB/b0903</td>
<td>0.75</td>
<td>96.86</td>
<td>0.001</td>
<td>18.81</td>
<td>97.2</td>
<td>8.93</td>
<td>30.62</td>
</tr>
<tr>
<td>pflC/b3952</td>
<td>0.75</td>
<td>96.86</td>
<td>0.001</td>
<td>18.81</td>
<td>97.2</td>
<td>8.93</td>
<td>30.62</td>
</tr>
<tr>
<td>pflD/b3951</td>
<td>0.75</td>
<td>96.86</td>
<td>0.001</td>
<td>18.81</td>
<td>97.2</td>
<td>8.93</td>
<td>30.62</td>
</tr>
</tbody>
</table>

Fig. 2. Pathways involved in micro-aerobic utilization of glucose by *E. coli* (WT Orth Model) to produce optically D-lactate and its constructed mutant strains (partially adopted from (Zhu and Shimizu, 2004; Zhu et al., 2007; Mienda et al., 2014a; Mienda and Shamsir, 2014). The pathways along with the deleted competing gene(s) are shown. The red colour enzymes represent the pathways that were inactivated via gene knockout. The sign “X” indicates gene knockouts. The knockout genes encode for Pyruvate formate lyase (pflA, pflB, pflC and pflD).
It is well known that in *E. coli*, PFL has several genes classes with corresponding subunits, primarily designated as *pflA*, *pflB*, *pflC* and *pflD* as mentioned earlier. It was also previously reported that PFL is a homodimeric protein with two subunits, PFL activating enzyme I and formate acetyltransferase I, are encoded by *pflA* and *pflB* genes respectively (Zhu and Shimizu, 2004). These two genes were believed to have constituted the *E. coli* PFL operon together with the probable formate transporter gene *focA* and anaerobically regulated promoters (Sawers, 1995). While the other two genes (*pflC* and *pflD*) probably code for PFL activating enzyme II and formate acetyltransferase II respectively (Zhu and Shimizu, 2004). On the bases of these findings *pflC* and *pflD* are not involved in the *E. coli* PFL operon as reported elsewhere (Zhu and Shimizu, 2004). Maximum uptake rates for glucose were set to be 20 mmol gDW⁻¹ h⁻¹ and the corresponding Oxygen uptake rate was 5 mmol gDW⁻¹ h⁻¹ for micro-aerobic simulation. Furthermore, the production of acetate and formate were reduced to nearly 97% of the wild-type model in all the mutant models examined (Table 1). This might be attributed to the deletion of pfl genes that are directly involved in acetate and formic acid formation (Fig. 2). Although acetate, formate and ethanol were also produced, indicating that other alternative pathways such as the use of pyruvate dehydrogenase complex (PDHc) may have been activated (Table 1 and Fig. 2). This is because it was previously reported that *E. coli* PDHc usually become activated when there is limited oxygen concentration or aerobic condition is established (Svetlana Alexeeva and Mattos, 2000) (Fig. 2).

**Conclusion**

The study informed other studies by demonstrating that the genome scale metabolic reconstruction of *E. coli* iJO1366 (Orth et al., 2011) in combination with the OptFlux software platform (Rocha et al., 2010) using ROOM (Shlomi et al., 2005) can prospectively and effectively predict metabolic engineering targets by *in silico* deletion of *pflA/b0902, pflB/b0903, pflC/b3952* and *pflD/b3951* for increased optically pure D-lactate production in *E. coli*. This would guide future experimental studies by accurately predicting the target for engineering using the OptFlux software platform. In addition, the behavior of the *E. coli* perturbed cells could be well understood, creating a path to model-guided novel biological insight for the production of value added chemicals via microbial cell factories.

**Acknowledgement**

We would like to thank Pablo Carbonell, Sara Careia and Paulo Maia for providing insight on the use of the OptFlux software platform.

**References**


Mienda BS, Shamsir MS, Shehu I, Deb A, Galadima IA (2014b). In silico metabolic engineering interventions of Escherichia coli for enhanced ethanol production, based on gene knockout simulation. IIOABJ 5: 16-23.


Biology 4: 45.


